

# **BAB I**

## **PENDAHULUAN**

### **A. Latar Belakang Masalah**

Evaluasi merupakan bagian dari sistem manajemen yaitu perencanaan, organisasi, pelaksanaan, monitoring dan evaluasi. Tanpa evaluasi, maka tidak akan diketahui bagaimana kondisi objek evaluasi tersebut dalam rancangan, pelaksanaan serta hasilnya. Istilah evaluasi sudah menjadi kosa kata dalam bahasa Indonesia, akan tetapi kata ini adalah kata serapan dari bahasa Inggris yaitu “*evaluation*” yang berarti penilaian atau penaksiran (Echols dan Shadily, 2000: 220).

Penelitian evaluatif adalah sebuah kegiatan yang bertujuan untuk mengetahui kinerja sebuah transformasi pembelajaran. Transformasi sendiri terdiri dari dua kata, yaitu (1) trans dan (2) formasi. Trans artinya proses perubahan, sedangkan formasi dari kata form, artinya bentuk. Jadi arti dari keseluruhan transformasi adalah perubahan bentuk. Makna dari proses pembelajaran adalah upaya untuk mengubah siswa yang semula masih dalam keadaan belum mengetahui ilmu yang diberikan oleh guru, sesudah melalui proses pembelajaran diharapkan mengetahui dengan baik. Penelitian evaluatif yang mengarahkan perhatiannya kepada proses pembelajaran, pertama berusaha

untuk mengetahui seberapa baik siswa telah menguasai materi pembelajaran yang diberikan oleh guru ( Arikunto, 2013: 41)

Perubahan kurikulum dari waktu ke waktu membuat proses evaluasi dalam dunia pendidikan mengalami perubahan waktu pelaksanaannya, dari kurikulum yang menggunakan sistem evaluasi catur wulan sampai dengan sekarang yang menggunakan sistem evaluasi per semester. Dengan diberlakukannya Kurikulum Tingkat Satuan Pendidikan (KTSP) yang merupakan pengembangan dari kurikulum 2004 yaitu Kurikulum Berbasis Kompetensi (KBK) maka program pengajaran dalam satu tahun dibagi menjadi dua semester sehingga untuk mengukur tingkat keberhasilan pengajaran dilakukan evaluasi tiap semester. Kegiatan evaluasi mempunyai peranan yang penting dalam pendidikan, begitu pula dalam proses pembelajaran karena dengan evaluasi dapat diketahui hasil dari kegiatan pembelajaran yang telah dilaksanakan dan dari hasil tersebut dapat ditentukan tindak lanjut yang akan dilakukan (Ariani, 2006: 73-74).

Evaluasi merupakan proses penentuan sampai sejauh mana kemampuan yang dapat dicapai oleh siswa dalam proses pembelajaran yang sebelumnya telah ditentukan besaran patokan nilainya, sehingga dapat dilakukan penilaian (*value judgment*). Proses evaluasi seperti yang telah kita tahu dapat dilakukan dengan kegiatan tes maupun non-tes. Alat evaluasi yang digunakan dalam penelitian ini adalah tes objektif pilihan ganda, karena bentuk ini dapat mencakup banyak materi pelajaran, penskoran objektif dapat dikoreksi siapa

saja dengan mudah, tingkat berpikir yang terlibat bisa dari tingkat pengetahuan sampai tingkat sintesis dan analisis (Mardapi, 2008: 60).

Tes dikatakan baik sebagai alat ukur apabila memenuhi persyaratan tes, yaitu memiliki: (1) validitas, (2) reliabilitas, (3) objektivitas, (4) praktisibilitas dan (5) ekonomis. Sebuah tes dikatakan valid apabila tes itu dapat tepat mengukur apa yang hendak diukur. Tes dikatakan reliable apabila memberikan hasil yang tepat apabila diteskan berkali-kali. Susunan tes dikatakan objektif apabila dalam melaksanakan tes itu tidak ada faktor subjektif yang mempengaruhi. Sebuah tes dikatakan memiliki praktisibilitas tinggi apabila tes tersebut bersifat praktis yaitu mudah dilaksanakan, mudah pemeriksaannya dan dilengkapi petunjuk-petunjuk yang jelas. Sedangkan persyaratan ekonomis artinya bahwa pelaksanaan tes tersebut tidak membutuhkan biaya yang mahal, tenaga yang banyak dan waktu yang lama (Arikunto, 2010: 23).

Seperti yang telah sebutkan di atas dengan berlakunya kurikulum tingkat satuan pendidikan (KTSP) guru diberi keleluasaan dalam melakukan penilaian mulai dari perencanaan sampai pelaksanaan, terutama dalam menyusun soal tes. Baik tidaknya soal tes sangat ditentukan oleh kemampuan guru dalam menyusun soal. Bahkan biasanya lebih tinggi lagi yaitu tingkat karisidenan, agar evaluasi yang dilakukan melalui penilaian dengan menggunakan tes sesuai dengan yang diharapkan maka diperlukan adanya peninjauan kembali terhadap pelaksanaan evaluasi yang telah dilaksanakan tersebut (Mardapi, 2008: 60).

Sekolah MTs Guppi 03 merupakan sekolah madrasah yang mengedepankan bidang agama, meskipun demikian sekolah ini juga terdapat mata pelajaran umum, seperti Matematika, Ilmu Pengetahuan Sosial, Ilmu Pengetahuan Alam dan lain sebagainya. Hasil observasi yang telah dilakukan di MTs Guppi 03 Belanga masih banyak kekurangan dalam sarana pembelajaran, sehingga menyebabkan siswa hanya terpaku pada buku paket dan media gambar pembelajaran. Dari hasil ulangan harian yang dilakukan pada sekolah tersebut masih ada siswa yang mendapatkan nilai rendah. Hasil nilai ulangan harian jika dirata-ratakan mendapatkan nilai 69.

Dengan hasil tersebut menuntut guru untuk lebih giat dalam memberikan pemahaman tentang materi yang akan disampaikan. Agar siswa lebih dapat meningkatkan nilai hasil belajar. Pada dasarnya siswa lebih cenderung pasif dalam belajar, sehingga menuntut guru untuk lebih giat membangun semangat siswa dalam melaksanakan kegiatan belajar mengajar. Meskipun telah diberikan motivasi pada siswa tetapi pada kenyataannya masih belum dapat meningkatkan semangat siswa dalam pembelajaran. Hal tersebut dibuktikan pada hasil dari nilai ulangan harian siswa rata-rata  $\leq 69$ , masih dibawah nilai KKM. Karena nilai KKM yang telah ditentukan dari pihak Guru Bidang Studi  $\geq 73$ .

Terdapat beberapa penelitian yang telah dilakukan sebelumnya yang relevan terhadap penelitian ini, antara lain: Hasil penelitian Arrizqi (2010 : 39) yaitu analisis UAS Biologi Kelas XI di Kabupaten Tegal pada soal pilihan ganda menunjukkan rata-rata persentase tingkat kesukaran mudah 29%, sedang 44%

dan sulit 26%; reliabilitassebesar 0,467; daya beda soal 80%; efektivitas pengecoh sebanyak 49% kurang berfungsi. Sedangkan analisis secara kualitatif menunjukkan bahwa 33 butir sesuai (aspek materi dan bahasa), tapi aspek konstruksi perlu direvisi.

Penelitian yang dilakukan oleh Afiyana (2010: 38) mengenai analisis soal latihan UN IPA diKabupaten Batang, menunjukkan kualitas soal dalam kategori kurang baik. Karena soal didominasi oleh tingkat kesukaran soal yang tinggi.

Hasil penelitian Anita Puspita Handayani (2012: 45), berdasarkan data angket validasi oleh tim ahli, diketahui bahwa instrumen tes pilihan ganda distraktor bermakna yang dikembangkan dinyatakan layak dalam ranah materi, konstruksi, dan bahasa. Setelah uji coba terbatas, data dianalisis untuk mengidentifikasi karakteristik konsep siswa, meliputi: analisis tiap indikator soal, analisis level tiap butir soal, analisis profil siswa, dan analisis profil kelas. Secara keseluruhan, analisis profil kelas menunjukkan bahwa siswa yang memahami konsep dengan benar pada materi getaran masih tergolong rendah, yaitu 29,68%, sedangkan 70,32% mengalami mis konsepsi.

Sumarsono (2000 : 40) melalui hasil Analisis Kualitas Soal Pendalaman Materi Bahasa Indonesia SLTP di Kabupaten Sleman menyatakan bahwa validitas soal Pendalaman Materi Bahasa Indonesia SLTP di Kabupaten Sleman tergolong baik. Reliabilitas soal tersebut termasuk rendah (0.09). Dilihat dari indeks kesulitan soal secara keseluruhan termasuk rendah. Ditinjau dari indeks daya pembeda secara keseluruhan soal kurang baik.

Analisis distraktor pada 70 soal pendalaman materi Bahasa Indonesia kelas III SLTP di Kabupaten Sleman Tahun Ajaran 1998/1999 secara keseluruhan tergolong baik.

Berdasarkan dari latar belakang tersebut di atas, maka peneliti tertarik melakukan penelitian tentang: Identifikasi Butir Soal IPA Pada Guru MTs Guppi 03 Belanga (Studi pada Butir Soal Ujian Tengah Semester Ganjil Siswa Kelas IX MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016)

#### **B. Rumusan Masalah**

Berdasarkan dari latar belakang tersebut di atas, maka dapat dirumuskan masalah penelitian yaitu:

1. Apakah butir soal Ulangan Tengah Semester Ganjil IPA kelas IX pada guru MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016 memiliki validitas tinggi?
2. Apakah butir soal Ulangan Tengah Semester Ganjil IPA kelas IX pada guru MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016 reliabel?
3. Bagaimanakah tingkat kesukaran butir soal Tengah Akhir Semester Ganjil IPA kelas IX pada guru MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016 tinggi?
4. Apakah daya beda butir soal Ulangan Tengah Semester Ganjil IPA kelas IX pada guru MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016 kuat?

5. Apakah pengecoh butir soal Ulangan Tengah Semester Ganjil IPA kelas IX pada guru MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016 berfungsi?

### **C. Tujuan Penelitian**

Penelitian ini bertujuan untuk menjawab permasalahan penelitian yakni untuk mengetahui:

1. Validitas soal Ulangan Tengah Semester Ganjil IPA kelas IX pada guru MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016
2. Reliabilitas soal Ulangan Tengah Semester GanjiIPA kelas IX pada guru MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016
3. Tingkat kesukaran soal Ulangan Tengah Semester Ganjil IPA kelas IX pada guru MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016
4. Daya beda soal Ulangan Tengah Semester Ganjil IPA kelas IX pada guru MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016
5. Efektivitas pengecoh butir soal Ulangan Tengah Semester Ganjil IPA kelas IX pada guru MTs Guppi 03 Belanga Lampung Selatan Tahun Pelajaran 2015/2016

#### **D. Manfaat Penelitian**

Hasil penelitian ini diharapkan dapat menjadi bahan masukan:

1. Bagi peneliti, yaitu untuk meningkatkan pengetahuan dan pemahaman sebagai calon guru tentang pembuatan soal yang baik dan benar sesuai dengan aturan yang telah ada,
2. Bagi siswa, yaitu agar siswa dapat memahami soal yang akan diberikan oleh guru serta dapat mengerjakannya dengan tepat dan benar,
3. Bagi Guru Bidang Studi, yaitu diharapkan dengan adanya identifikasi soal guru bidang studi lebih dapat meningkatkan kualitas soal,
4. Bagi sekolah, yaitu memberikan sumbangan pemikiran untuk meningkatkan pembuatan soal biologi di sekolah melalui identifikasi butir soal.

#### **E. Ruang Lingkup Penelitian**

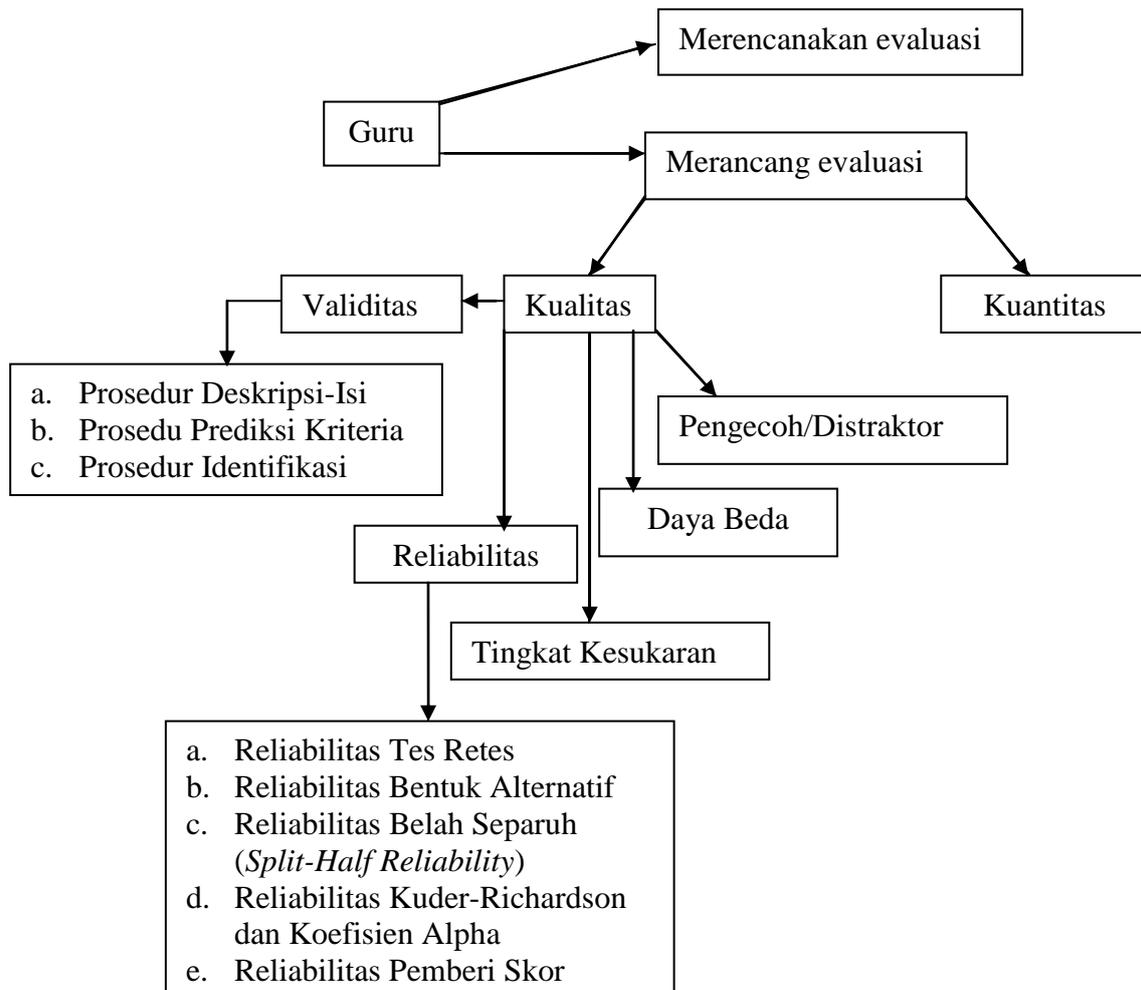
Ruang lingkup penelitian ini adalah:

1. Objek kajian pada penelitian ini adalah butir soal ujian tengah semester ganjil siswa kelas IX MTs Guppi 03 Belanga Lampung Selatan 2015/2016,
2. Hasil yang diukur dalam penelitian ini adalah validitas, reliabilitas, tingkat kesukaran, daya beda, dan keefektifan distraktor/pengecoh,
3. Sampel penelitian ini adalah seluruh jawaban tes ujian tengah semester ganjil bidang studi IPA dari siswa kelas IX MTs Guppi 03 Belanga Lampung Selatan.

## **F. Kerangka Pikir**

Banyak sekali cara yang dapat ditempuh untuk melakukan penilaian terhadap hasil belajar siswa, baik melalui cara tes maupun non-tes, usaha-usaha yang telah dilakukan dengan jalan tes yang digunakan untuk mengukur tingkat kecerdasan siswa tidak akan berfungsi dengan baik mana kala soal-soal yang diujikan tidak memiliki kualitas yang baik. Baik tidaknya kualitas sebuah soal dapat dilihat dari tingkat kevalidan, reliabilitas soal, daya beda soal, tingkat kesukaran soal dan tingkat efektifitas distraktor/pengecoh.

Oleh karena itu agar suatu tes dapat mengukur dengan baik maka harus memiliki kualitas yang baik juga, untuk mengetahui baik tidaknya sebuah instrument tes maka harus melalui tahap peninjauan kembali yaitu dengan menganalisis soal-soal yang telah ada, sehingga dapat dilakukan revisi apabila terdapat kekurangan. Tes dikatakan baik sebagai alat ukur apabila memenuhi persyaratan tes, yaitu memiliki: (1) validitas, (2) reliabilitas, (3) objektivitas, (4) praktisibilitas dan (5) ekonomis



Gambar 1. Perencanaan Evaluasi

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **A. Evaluasi Pendidikan**

Evaluasi merupakan salah satu sarana penting dalam meraih tujuan belajar mengajar. Guru sebagai pengelola kegiatan belajar mengajar dapat mengetahui kemampuan yang dimiliki siswa, ketepatan metode mengajar yang digunakan, dan keberhasilan siswa dalam meraih tujuan pembelajaran yang telah ditetapkan melalui kegiatan evaluasi. Guru dapat mengambil keputusan secara tepat dengan informasi ini mengenai langkah apa yang harus dilakukan selanjutnya. Informasi tersebut juga dapat memberikan motivasi kepada siswa untuk berprestasi lebih baik.

Ujian *tryout* pada hakikatnya merupakan evaluasi hasil belajar yang dilaksanakan oleh lembaga pendidikan sebelum menghadapi ujian nasional (UN). Hal-hal yang berkaitan dengan evaluasi belajar, seperti: pengertian, tujuan dan fungsi, prinsip evaluasi, dan jenis evaluasi akan diuraikan sebagai berikut.

Ada bermacam-macam pengertian evaluasi pendidikan yang dikemukakan oleh para ahli. Menurut Ratumanan (2003: 1), evaluasi dapat dinyatakan sebagai suatu proses sistematis dalam menentukan tingkat pencapaian tujuan instruksional. Ralp Tyler ( dalam Arikunto, 2011: 3) mengatakan bahwa

“Evaluasi merupakan sebuah proses pengumpulan data untuk menentukan sejauh mana, dalam hal apa, dan bagaimana tujuan pendidikan sudah tercapai. Jika belum, bagaimana yang belum dan apa sebabnya”. Definisi yang lebih luas dikemukakan oleh dua orang ahli, yakni Cronbach dan Stufflebeam. Tambahan definisi tersebut adalah bahwa proses evaluasi bukan sekedar mengukur sejauh mana tujuan tercapai, tetapi digunakan untuk membuat keputusan. Jika evaluasi dikaitkan dengan pendidikan maka evaluasi pendidikan memiliki dua konsep pengertian. Hal ini sejalan dengan pendapat Sudijono (2006: 2) bahwa evaluasi pendidikan adalah: (1) Proses/kegiatan untuk menentukan kemajuan pendidikan, dibandingkan dengan tujuan yang telah ditentukan; (2) Usaha untuk memperoleh informasi berupa umpan balik (*feed back*) bagi penyempurnaan pendidikan.

Kesimpulan yang dapat diambil melalui beberapa konsep pengertian di atas, evaluasi pendidikan adalah suatu proses sistematis yang mengukur, menelaah, menafsirkan, dan mempertimbangkan sekaligus memberikan umpan balik (*feed back*) untuk mengetahui tingkat pencapaian terhadap tujuan pembelajaran yang telah ditetapkan serta digunakan sebagai informasi untuk membuat keputusan.

Kegiatan evaluasi juga mempunyai tujuan khusus dalam bidang pendidikan, yaitu: (a) untuk merangsang kegiatan peserta didik dalam menempuh program pendidikan, dan (b) untuk menemukan faktor-faktor penyebab keberhasilan dan ketidak berhasilan peserta didik dalam mengikuti program pendidikan, sehingga dapat dicari dan ditemukan jalan keluar atau cara-cara perbaikannya.

Sudijono (2006: 7) menjelaskan bahwa secara umum ada tiga fungsi evaluasi, yaitu untuk: (a) mengukur kemajuan, (b) menunjang penyusunan rencana, dan (c) memperbaiki atau melakukan penyempurnaan kembali. Sudijono juga menambahkan, bahwa selain memiliki fungsi secara umum evaluasi juga memiliki fungsi secara khusus. Adapun fungsi evaluasi secara khusus dalam bidang pendidikan dapat ditinjau dari tiga segi, yaitu: (a) segi psikologi, (b) segi didaktik, dan (c) segi administratif.

Evaluasi pendidikan secara psikologi akan memberikan petunjuk untuk mengenal kemampuan dan status dirinya di antara kelompok atau kelasnya. Siswa akan mengetahui apakah dirinya termasuk berkemampuan tinggi, rata-rata, atau rendah. Apabila hal tersebut dapat dicapai maka diharapkan evaluasi pendidikan akan dapat memberikan dorongan kepada siswa untuk memperbaiki, meningkatkan, dan mempertahankan prestasinya.

Evaluasi pendidikan bagi pendidik secara didaktik, setidaknya memiliki lima macam fungsi, yaitu: (1) memberikan landasan untuk menilai hasil usaha (prestasi) yang telah dicapai oleh peserta didik, (2) memberikan informasi yang sangat berguna untuk mengetahui posisi masing-masing siswa diantara kelompoknya, (3) memberikan bahan penting untuk memilih dan kemudian menetapkan status peserta didik, (4) memberikan pedoman untuk mencari dan menemukan jalan keluar bagi siswa yang memerlukannya, dan (5) memberikan petunjuk sejauh mana tujuan program pengajaran yang telah ditentukan telah dicapai.

Evaluasi pendidikan secara administrasi setidaknya memiliki tiga macam fungsi yaitu: (1) memberikan laporan mengenai kemajuan dan perkembangan siswa yang telah mengikuti kegiatan pendidikan dan pelatihan dalam jangka waktu tertentu, (2) memberikan bahan-bahan keterangan (data) untuk keperluan pengambilan keputusan, dan (3) memberikan gambaran mengenai hasil-hasil yang telah dicapai dalam proses pembelajaran.

Mengingat pentingnya evaluasi dalam menentukan kualitas pendidikan, maka upaya merencanakan dan melaksanakan evaluasi hendaknya memperhatikan beberapa prinsip.

Menurut Daryanto (2005: 19-21), terdapat beberapa prinsip yang perlu diperhatikan dalam melakukan evaluasi, yaitu keterpaduan, keterampilan siswa, koherensi, pedagogis, dan akuntabilitas.

#### 1) Keterpaduan

Tujuan instruksional, materi, metode pengajaran, serta evaluasi merupakan tiga kesatuan erpadu yang tidak boleh dipisahkan. Oleh karena itu, perencanaan evaluasi harus sudah ditetapkan pada waktu menyusun suatu pengajaran sehingga dapat disesuaikan secara harmonis dengan tujuan instruksional dan materi pengajaran yang hendak disajikan.

#### 2) Keterlibatan Siswa

Untuk mengetahui sejauh mana siswa berhasil dalam kegiatan belajar mengajar yang dijalani secara aktif, siswa membutuhkan evaluasi. Penyajian evaluasi oleh guru merupakan upaya guru untuk memenuhi kebutuhan siswa

akan informasi mengenai kemajuannya dalam program belajar mengajar.

Siswa akan merasa kecewa apabila usahanya tidak dievaluasi.

### 3) Koherensi

Prinsip evaluasi dimaksudkan evaluasi harus berkaitan dengan materi pengajaran yang sudah disajikan dan sesuai dengan ranah kemampuan yang hendak diukur.

### 4) Pedagogis

Evaluasi dan hasil hendaknya dapat dipakai sebagai alat motivasi untuk siswa dalam kegiatan belajarnya.

### 5) Akuntabilitas

Evaluasi dan hasilnya dapat dipakai sebagai laporan pertanggung jawaban kepada pihak-pihak yang berkepentingan dengan pendidikan sehingga dapat diketahui sejauh mana keberhasilan pembelajaran yang telah dilakukan.

Dalam merencanakan dan melakukan evaluasi pembelajaran, seorang guru hendaknya selalu berpegang pada prinsip-prinsip tersebut. Hal ini dimaksudkan agar guru dapat bertindak dan berusaha seobjektif mungkin dalam mengadakan evaluasi.

Menurut Daryanto (2005: 11-14) untuk masing-masing tindak lanjut yang dikehendaki dalam evaluasi diadakan tes yang disebut tes penempatan, tes formatif, tes diagnostik, dan tes sumatif. Tes penempatan dilaksanakan pada awal tahun pelajaran baru, sebagai alat ukur untuk mengetahui tingkat kemampuan yang telah dimiliki peserta didik. Dengan demikian, siswa dapat ditempatkan pada kelompok yang sesuai dengan tingkat pengetahuan yang telah dimilikinya.

Tes formatif dilaksanakan di tengah program pembelajaran untuk memantau kemajuan belajar siswa demi memberikan umpan balik, baik kepada siswa maupun kepada guru. Berdasarkan hasil tes tersebut dapat diketahui materi pelajaran apa yang belum dikuasai siswa sehingga guru harus mengupayakan perbaikannya.

Tes diagnostik digunakan untuk mendiagnosa kesalahan belajar siswa dan mengupayakan perbaikannya. Pada jenis ini, tes formatif terlebih dahulu disajikan untuk mengetahui ada tidaknya bagian mana yang belum dikuasai siswa, sehingga dapat dibuat butir-butir soal yang tingkat kesukarannya relatif rendah untuk mendekteksi. Tes sumatif diberikan pada akhir tahun ajaran untuk memberikan nilai sebagai dasar menentukan kelulusan atau pemberian sertifikat bagi siswa yang telah menyelesaikan pelajaran dengan baik. Ruang lingkup tes sumatif mencakup seluruh bahan yang telah disajikan sepanjang jenjang pendidikan.

## **B. Instrumen Evaluasi**

Dalam pengertian umum, alat adalah sesuatu yang dapat digunakan untuk mempermudah seseorang untuk melaksanakan tugas atau mencapai tujuan secara efektif dan efisien. Kata “alat” biasa juga disebut dengan istilah “instrumen”. Dengan demikian maka alat evaluasi juga dikenal dengan instrumen evaluasi (Arikunto, 2011: 25-26).

Kegiatan evaluasi tidak dapat dipisahkan dari suatu instrumen evaluasi yang digunakan untuk mendapatkan data evaluasi. Untuk mendapatkan data

evaluasi, evaluator menggunakan cara atau teknik. Cara tersebut dikenal dengan teknik evaluasi. Teknik evaluasi dapat dibedakan menjadi dua, yaitu teknik tes dan nontes. Kedua teknik tersebut digunakan untuk mencari data-data pengukuran tentang keberhasilan siswa dalam kegiatan belajar mengajar. Instrumen yang banyak digunakan dalam bidang pendidikan ialah tes. Tes adalah cara yang digunakan untuk melakukan pengukuran dalam bentuk tugas-tugas yang harus dikerjakan siswa agar mendapatkan data tentang nilai prestasi siswa tersebut untuk dibandingkan dengan nilai standar yang ditetapkan (Nur kencana dan Sumartana, 1983: 25). Bentuk tes dapat dibedakan menjadi dua, yaitu tes objektif dan tes subjektif.

Menurut Nurgiyantoro (2001: 11), tes subjektif digunakan agar siswa menunjukkan kemampuan dalam menerapkan pengetahuan, menganalisis, dan mengevaluasi informasi baru yang diharapkan kepadanya. Jawaban siswa terhadap tes subjektif menunjukkan kualitas cara berpikir siswa, aktifitas kognitif yang bukan semata-mata mengingat dan memahaminya saja.

Tes objektif adalah tes yang bersifat pasti, karena hanya ada satu kemungkinan jawaban yang benar. Sebagai alat pengukur hasil belajar siswa, tes ini sering dipilih karena dapat dikerjakan dengan cepat dan mencakup materi ajar yang luas, hanya memiliki satu jawaban yang pasti sehingga memudahkan korektor untuk mengoreksi, serta dapat dikoreksi secara cepat dengan hasil yang dapat dipercaya.

Menurut Arikunto (2011: 26) yang tergolong teknik nontes adalah skala bertingkat, kuesioner, daftar cocok, wawancara, pengamatan, dan riwayat hidup. Teknik nontes merupakan cara yang digunakan untuk memperoleh informasi yang berkaitan dengan tingkah laku kognitif. Non tes biasanya berisi pertanyaan atau pernyataan yang tidak memiliki jawaban benar atau salah secara mutlak.

Kegiatan evaluasi bentuk tes dan jenis tagihan yang digunakan hendaknya bervariasi tergantung pada jenis data yang ingin diperoleh. Dalam kaitannya dengan penelitian ini, jenis tagihan yang digunakan dalam ujian *tryout* berupa soal tes pilihan ganda. Materi soal yang diujikan harus disusun sesuai dengan kisi-kisi soal yang telah dirumuskan.

### **C. Acuan Norma dan Kriteria**

Acuan norma dan acuan kriteria menggunakan asumsi yang berbeda tentang kemampuan seseorang. Penafsiran hasil tes antara dua acuan ini berbeda hingga menghasilkan informasi yang berbeda maknanya. Tes acuan norma berasumsi bahwa kemampuan setiap orang itu berbeda dan dapat digambarkan menurut distribusi normal. Perbedaan ini harus ditunjukkan oleh hasil pengukuran, misalnya setelah mengikuti tes peserta didik dibandingkan dengan kelompoknya. Hal itu berarti standar yang diterapkan pada satu kelompok tidak dapat diterapkan pada kelompok lain. Standar yang dibuat berdasarkan hasil tes sebelumnya pun tidak dapat digunakan untuk hasil tes yang sekarang dan yang akan datang. Jadi, norma yang satu akan berbeda dengan norma yang lain. Kesemuanya bergantung

pada prestasi siswa pada kelompok yang bersangkutan (Nurgiyantoro, 2001: 404).

Berbeda dengan pendekatan acuan norma, pendekatan acuan kriteria berusaha menafsirkan hasil tes yang ditetapkan standar yang akan dipergunakan sebelum kegiatan pengajaran dilakukan (Nurgiyantoro, 2001: 397). Pada acuan kriteria seorang siswa dikatakan lulus apabila telah memenuhi tingkat penguasaan minimal pada bahan pelajaran. Konsekuensi dari acuan ini adalah adanya remidi dan pengayaan. Siswa yang belum memenuhi tingkat penguasaan minimal terhadap bahan pelajaran seperti yang diisyaratkan harus belajar lagi sampai kemampuannya mencapai standar yang telah ditetapkan. Sementara itu, bagi siswa yang telah mencapai standar diberi pelajaran tambahan yang disebut program pengayaan.

#### **D. Validitas**

Manfaat penilaian adalah untuk menentukan apakah sebuah tes telah mengukur apa yang hendak diukur dan apakah sebuah tes telah tepat digunakan untuk membuat suatu keputusan tentang pengambilan tes. Pada saat pengujian suatu soal, untuk mengambil keputusan tentang belajar siswa, harus dipastikan bahwa soal tersebut valid dan reliabel. Validitas soal dapat dilihat dari kisi-kisi soal, sedang reliabilitas soal baru dapat diketahui setelah dilakukan uji coba terhadap soal tersebut. Uji coba ini dilakukan terhadap sejumlah subjek yang bersifat tipikal dengan populasi yang akan dites (Nurgiyantoro, 2001: 135).

Validitas atau (kesahihan) tes dapat diartikan sebagai ketepatan dan kecermatan tes dalam menjalankan fungsi pengukurannya. Suatu tes mempunyai validitas tinggi apabila tes tersebut menjalankan fungsi ukurnya, atau memberikan hasil ukur yang sesuai dengan maksud dan tujuan diadakannya tes tersebut. Sebaliknya, suatu tes yang menghasilkan data tidak relevan dengan tujuan pengukuran dikatakan sebagai tes yang memiliki validitas rendah (Azwar,1987:146).

Menurut Arikunto (2011: 65-69) validitas sebuah tes dapat diketahui dari hasil pengalaman. Validitas tes tersebut dibagi menjadi dua kelompok, yaitu validitas logis dan validitas empiris. Validitas logis meliputi validitas isi (*content validity*) dan validitas konstruk (*construct validity*), sedangkan validitas empiris meliputi validitas “ada sekarang” atau konkruen (*concurrent validity*) dan validitas prediksi (*predictive validity*).

1) Validitas Isi (*content validity*)

Validitas isi didefinisikan sebagai kecocokan di antara isi alat ukur dengan sasaran ukur. Untuk keperluan pencocokan, biasanya isi sasaran ukur disusun dalam bentuk spesifikasi, yang meliputi bahan atau materi dan tujuan hasil belajar.

Sebuah tes dikatakan memiliki validitas isi apabila mengukur tujuan khusus tertentu yang sejajar dengan materi atau isi pelajaran yang diberikan. Oleh karena materi yang diajarkan tertera dalam kurikulum, maka validitas isi ini sering juga disebut dengan validitas kurikuler. Validitas isi ini diusahakan tercapai sejak saat penyusunan dengan cara merinci materi kurikulum

atau materi buku pelajaran.

2) Validitas Konstruksi (*construct validity*)

Arikunto mengatakan bahwa sebuah tes memiliki validitas konstruk apabila butir-butir soal yang membangun tes tersebut mengukur setiap aspek berpikir seperti yang disebutkan dalam tujuan instruksional. Konstruk dalam pengertian ini bukanlah susunan, tetapi merupakan rekaan psikologis, yaitu suatu rekaan yang dibuat oleh para ahli ilmu jiwa yang dengan suatu cara tertentu merinci isi jiwa atas beberapa aspek seperti: ingatan, pemahaman, aplikasi, dan seterusnya.

3) Validitas Konkruen (*concurrent validity*)

Validitas ini lebih umum dikenal dengan validitas empiris. Sebuah tes dikatakan memiliki validitas empiris jika hasilnya sesuai dengan pengalaman. Dalam membandingkan hasil sebuah tes maka diperlukan suatu alat pembanding, maka hasil tes merupakan sesuatu yang dibandingkan.

4) Validitas Prediksi (*predictive validity*)

Menurut Arikunto sebuah tes dikatakan memiliki validitas prediksi atau validitas ramalan apabila mempunyai kemampuan untuk meramalkan apa yang terjadi pada masa yang akan datang.

Untuk mengetahui apakah validitas isi telah dipenuhi atau belum, maka dilakukan telaah soal dengan cara menyesuaikan soal dengan kisi-kisi yang diacu. Dalam hal ini, digunakan *expert judgment* untuk menilai kesesuaian itu. Selain itu, untuk mengetahui validitas isi juga digunakan kartu telaah butir soal yang mencakup bidang kriteria penelaahan materi, konstruksi, dan

bahasa.

### **E. Reliabilitas**

Reliabilitas suatu tes adalah konsistensi dari suatu tes dalam mengukur apa yang seharusnya diukur sehingga pengukuran itu memberikan informasi yang dapat dipercaya. Pengertian itu disimpulkan dari pendapat dua ahli pengukuran berikutini. Arikunto (2011:86) menyatakan bahwa reliabilitas dalam pengukuran berhubungan dengan masalah kepercayaan. Menurut Suryabrata (2004: 29), reliabilitas menunjukkan sejauh mana hasil pengukuran dengan alat tersebut dapat dipercaya. Hasil pengukuran harus reliabel artinya harus memiliki tingkat konsistensi dan kemampuan.

Suatu alat tes dikatakan reliabel apabila alat tes tersebut dapat dipercaya, konsisten, atau tetap. Untuk membuktikan apakah suatu alat tes memiliki sifat tetap, perlu diadakan uji coba terhadap alat tes yang akan digunakan tersebut. Apabila sebuah tes diujicobakan lebih dari satu kali pada subjek yang sama dapat menghasilkan data yang kurang lebih sama, tes tersebut dikatakan dapat mengukur secara tetap.

Untuk mengukur reliabilitas soal dapat digunakan beberapa metode antara lain metode stabilitas, metode ekuavalensi, dan metode konsistensi internal (Surapranata, 2004: 90). Metode yang tepat digunakan pada penelitian ini adalah metode konsistensi internal. Metode konsistensi internal diperoleh dengan cara menganalisis data dari satu kali pengumpulan data (Widoyoko, 2009:147).

Nunnally dalam Surapranata (2004: 113) menyatakan bahwa metode ini didasarkan pada korelasi antar skor jawaban pada setiap butir tes. Ia juga menyatakan bahwa metode ini khususnya digunakan pada setiap butir-butir yang dikotomi seperti pilihan ganda.

Reliabilitas soal dapat diketahui dari nilai *Alpha Cronbach* yang dihasilkan dari analisis kuantitatif *MicroCat Iteman*. Menurut Nurgiyantoro (2004: 332) untuk tes-tes yang distandarkan, harga indeks reliabilitas itu paling tidak harus mencapai 0,85 sampai 0,90. Soal *try out* bersama mata pelajaran bahasa Indonesia ini disusun oleh tim MGMP (Musyawarah Guru Mata Pelajaran) sehingga koefisien reliabilitas yang relevan adalah sebesar 0,90 atau lebih dari 0,90. Semakin tinggi koefisien reliabilitas berarti semakin baik soal tersebut. Soal yang memiliki koefisien reliabilitas yang tinggi berarti kesalahan penyusunannya semakin kecil. Jumlah butir pertanyaan akan mempengaruhi indeks reliabilitas instrumen yang bersangkutan. Semakin banyak butir pertanyaan akan semakin tinggi indeks reliabilitasnya.

Reliabilitas soal dalam penelitian ini diuji dengan teknik reliabilitas *Alpha Cronbach*. Untuk membantu perhitungan digunakan komputer dengan program *Micro Cat Iteman* versi 3.00.

#### **F. Tingkat Kesukaran**

Arikunto (2011: 207) menyatakan bahwa “soal yang memiliki indeks kesulitan yang baik adalah soal yang tidak terlalu mudah tetapi tidak terlalu sukar”. Butir soal dikatakan mudah apabila sebagian besar siswa

dapat menjawab dengan benar dan dikatakan sulit apabila sebagian besar siswa tidak dapat menjawab dengan benar. Didalam istilah evaluasi, indeks kesulitan ini diberi simbol P (p besar), singkatan dari kata “proporsi”. Besarnya indeks kesulitan (harga P) dapat dilihat dari proporsi peserta tes yang menjawab benar terhadap butir soal.

Tingkat kesulitan suatu butir soal dinyatakan dengan sebuah indeks yang berkisar antara 0,0 sampai dengan 1,0. Soal dengan indeks kesulitan 0,0 menunjukkan bahwa butir soal itu terlalu sulit, sebaliknya soal dengan indeks 0,1 menunjukkan bahwa soal yang bersangkutan terlalu mudah.

Penentuan indeks kesulitan (harga P) yang baik disesuaikan dengan jenis dan tujuan tes yang digunakan. Perbedaan jenis dan fungsi tes menghendaki harga P yang berbeda pula. Terkait dengan ketentuan tersebut, belum ditentukan harga p yang dianggap baik untuk suatu tes, perlu diketahui harga P tersebut akan dikenakan pada tes apa.

Indeks kesulitan yang sesuai untuk tes penempatan adalah sebesar 0,30-0,70; tes diagnosti dan tes sumatif sebesar 0,25–0,75 Gofur (dalam Khasanah, 2006: 28-29). Penelitian mengenai analisis soal *tryout* ini tergolong tes sumatif sehingga indeks kesulitan yang digunakan sebesar 0,25-0,75. Pada analisis *Iteman*, indeks kesulitan butir soal ditunjukkan pada kolom *Proportion Correct*. Kriteria indeks kesulitan soal berkategori baik berada pada posisi 0,25-0,75, indeks kesulitan pada posisi  $\leq 0,25$  berkategori sulit, dan indeks kesulitan  $\geq 0,75$  tergolong dalam kategori mudah. Butir soal yang baik adalah butir soal yang memiliki indeks kesulitan

sedang. Sedangkan butir soal yang memiliki indeks kesulitan mudah dan sulit, maka butir soal tersebut tergolong jelek sehingga perlu direvisi atau diganti.

### **G. Daya Beda**

Daya beda mengkaji butir-butir soal dengan tujuan untuk mengetahui kesanggupan soal dalam membedakan antara siswa yang tergolong mampu dengan siswa yang tergolong kurang mampu (Nana, 2005: 141).

Butir soal dinyatakan layak jika indeks daya bedanya baik. Indeks daya beda dapat berkisar antara -1,00 sampai dengan +1,00, tetapi jika indeks mendekati nol negatif maka dinyatakan tidak layak. Beberapa ahli mengatakan bahwa sebuah butir soal dikatakan layak apabila memiliki indeks daya beda sebesar 0,25 dan ada pula yang menyatakan 0,30. Penelitian ini melibatkan peserta ujian yang banyak, maka ditetapkan indeks daya beda yang layak minimum 0,30. Pada analisis iteman, indeks daya beda ditunjukkan oleh *Point Biser*.

Indeks daya beda diabaikan karena acuan kriteria tidak menentukan perbedaan antara yang bodoh dan yang pintar. Dalam acuan kriteria pada prinsipnya semua siswa memiliki kemampuan yang sama dan bisa belajar apa saja, hanya waktu yang diperlukan untuk mencapai kemampuan tersebut berbeda-beda. Dalam Kurikulum Tingkat Satuan Pendidikan, indeks daya beda juga tidak digunakan, karena indeks daya beda akan dideskripsikan sebagai hasil penemuan analisis butir soal.

## H. Keefektifan Distraktor

Tujuan analisis distraktor adalah untuk mengetahui seberapa banyak siswa menjawab benar sesuai dengan kunci jawaban dan seberapa banyak yang memilih distraktor atau pengecoh. Menurut Suyata (dalam Nurwanti, 2008: 28), suatu pengecoh dikatakan berfungsi dengan baik jika paling sedikit dipilih oleh 2% peserta tes.

Dipihak lain, Azwar (2006: 151-152) menyatakan bahwa alternatif yang merupakan distraktor yang baik harus memiliki koefisien korelasi yang negatif, karena hal itu mengidentifikasikan bahwa pemilihannya memang berasal dari peserta didik yang kurang cakap. Dalam hal ini, analisis distraktor hanya dapat diterapkan pada tes pilihan ganda (tes objektif). Butir soal yang baik adalah butir soal yang kunci jawabannya positif, distraktornya negatif. Pada analisis *Iteman*, keefektifan distraktor ditunjukkan pada kolom *Proportion Endorsing*.

## **BAB III**

### **METODE PENELITIAN**

#### **A. Penelitian**

Metode yang dipakai dalam penelitian ini adalah metode deskriptif karena penelitian ini bertujuan untuk mengetahui identifikasi butir soal IPA Siswa Kelas IX MTs Guppi 03 Belanga Lampung Selatan Semester Ganjil Tahun Pelajaran 2015/2016.

#### **B. Penentuan Objek Penelitian**

##### **a. Objek Penelitian**

Objek dalam penelitian ini adalah semua butir soal pilihan jamak Ujian Tengah Semester ganjil bidang studi IPA kelas IX MTs Guppi 03 Belanga Lampung Selatan. Beberapa hal yang diteliti dalam penelitian ini meliputi validitas, reliabilitas, daya beda soal, tingkat kesukaran soal, dan efektifitas distraktor.

##### **b. Sampel**

Teknik sampling yang digunakan adalah *purposive Sampling*, dimana sampel dalam penelitian ini adalah seluruh jawaban tes Ujian Tengah Semester ganjil bidang studi IPA dari siswa kelas IX MTs Guppi 03 Belanga Lampung

Selatan yang akan ditentukan kelasnya. Sampel diperoleh dengan menggunakan sistem 27% yaitu dengan cara:

1. Seluruh tes hasil belajar siswa yang akan diambil dari siswa MTs GUPPI 03 Belanga diurutkan berdasarkan skor dari yang tinggi sampai rendah.
2. Mengambil 27% skor teratas sebagai kelompok atas dan 27% skor terendah sebagai kelompok bawah. Sampel yang diambil dari kelompok atas dan kelompok bawah dijumlah yang nantinya akan digunakan untuk perhitungan daya beda soal, tingkat kesukaran soal dan efektivitas distraktor.

### **C. Pengumpulan Data**

Metode pengumpulan data dalam penelitian ini menggunakan metode dokumentasi dan observasi yaitu untuk mendapatkan seperangkat soal dan jawaban, daftar nama-nama siswa kelas IX MTs Guppi 03 Belanga Lampung Selatan, standar isi mata pelajaran IPA serta kisi-kisi penulisan soal.

### **D. Analisis Data**

Data yang dianalisis berupa hasil tes belajar siswa yang diambil dari siswa kelas IX MTs Guppi 03 Belanga Lampung Selatan. Dari hasil analisis yang diperoleh ditabulasikan dan didiskripsikan sesuai dengan kriteria yang telah ditentukan.

Analisis soal dilakukan dengan mengukur beberapa variabel yaitu validitas, reliabilitas, daya beda soal, tingkat kesukaran soal, dan efektifitas

distraktor/pengecoh. Cara pengukuran masing-masing variabel tersebut adalah sebagai berikut:

#### a. Validitas

Validitas yaitu ketepatan mengukur yang dimiliki oleh sebutir item (yang merupakan bagian tak terpisahkan dari tes sebagai suatu totalitas), dalam mengukur apa yang seharusnya diukur lewat butir item tersebut.

Sebutir item dapat dikatakan telah memiliki validitas yang tinggi atau dapat dikatakan valid, jika skor-skor pada butir item yang bersangkutan memiliki kesesuaian atau kesejajaran arah dengan skor totalnya, atau dengan bahasa statistik: Ada korelasi positif yang signifikan antara skor item dengan skor totalnya. Skor total disini berkedudukan sebagai variabel terikat (*dependent variable*), sedangkan skor item berkedudukan sebagai variabel bebasnya (*independent variable*). Untuk sampai pada kesimpulan bahwa item-item yang ingin diketahui validitasnya, yaitu valid ataukah tidak, kita dapat menggunakan teknik korelasi sebagai teknik analisisnya. Sebutir item dapat dinyatakan valid, apabila skor item yang bersangkutan terbukti mempunyai korelasi positif yang signifikan dengan skor totalnya.

$$r_{pbi} = \frac{M_p - M_t}{SD_t} \sqrt{\frac{p}{q}}$$

dimana :

$r_{pbi}$  = Koefisien korelasi point biserial yang melambangkan kekuatan korelasi antara variabel I dengan variabel II, yang dalam hal ini dianggap sebagai Koefisien Validitas Item.

$M_p$  = Skor rata-rata hitung yang dimiliki oleh *testee*, yang untuk butir item yang bersangkutan telah dijawab dengan betul.

$M_t$  = Skor rata-rata dari skor total.

$SD_t$  = Deviasi standar dari skor total.

$P$  = Proporsi *testee* yang menjawab betul terhadap butir item yang sedang diuji validitas itemnya.

$q$  = Proporsi *testee* yang menjawab salah terhadap butir item yang sedang diuji validitas itemnya.

(Arikunto, 2013: 85)

Tabel 1. Klasifikasi Interpretasi Koefisien Validitas

Besarnya $r_{hitung}$	Interpretasi
$0,90 \leq r_{hitung} \leq 1,00$	Validitas sangat tinggi
$0,70 \leq r_{hitung} < 0,90$	Validitas tinggi
$0,40 \leq r_{hitung} < 0,70$	Validitas sedang
$0,20 \leq r_{hitung} < 0,40$	Validitas rendah
$0,00 \leq r_{hitung} < 0,20$	Validitas sangat rendah
$r_{hitung} < 0,00$	Tidak valid

(Arikunto, 2013: 90)

Adapun hasil analisis uji instrumen mengenai validitas tiap butir soal seperti pada tabel berikut ini :

Tabel 2. Hasil Perhitungan Nilai Validitas Tiap Butir Soal Uraian

Nomor soal	Nilai Koefisien validitas	Klasifikasi indeks Validitas
1	0,55	Sedang
2	0,58	Sedang
3	0,56	Sedang
4	0,57	Sedang
5	0,74	Tinggi
6	0,74	Tinggi

(Arikunto, 2013: 99)

Berdasarkan klasifikasi koefisien validitas pada Tabel 3.1 dapat disimpulkan bahwa instrumen penelitian ini diinterpretasikan sebagai soal yang mempunyai validitas sedang dan validitas tinggi.

### b. Reliabilitas soal

Reliabilitas adalah proporsi dari varian dengan varian yang sesungguhnya.

Reliabilitas soal diukur dengan menggunakan rumus Kuder Richardson.

$$r_{KR20} = \frac{k}{k-1} \left( 1 - \frac{\sum pq}{s^2} \right)$$

Keterangan :

$r_{11}$  : reliabilitas tes  
 $k$  : banyaknya butir soal  
 $M$  : skor rata-rata  
 $V_t$  : Varians total  
 (Suherman, 2003: 123)

$V_t$  dicari dengan menggunakan rumus:

$$s^2 = \frac{\sum X^2}{N} - \bar{X}^2$$

Dimana :

$N_p$  : jumlah siswa yang menjawab benar  
 $N$  : jumlah siswa  
 (Suherman, 2003: 134)

Apabila  $r_{11} > r_{tabel}$  maka tes tersebut reliabel. Untuk menentukan tinggi rendahnya koefisien reliabilitas ( $r$ ) dapat ditentukan berdasarkan kriteria sebagai berikut, apabila harga  $r$ :

Tabel 3. Klasifikasi Koefisien Reliabilitas

Besar $r_{11}$	Interpretasi
$r_{11} \leq 0,20$	Reliabilitas sangat rendah
$0,20 \leq r_{11} < 0,40$	Reliabilitas rendah
$0,40 \leq r_{11} < 0,70$	Reliabilitas sedang
$0,70 \leq r_{11} < 0,90$	Reliabilitas tinggi
$0,90 \leq r_{11} < 1,00$	Reliabilitas sangat tinggi

(Suherman, 2003:145)

### c. Daya Beda Soal

Daya pembeda item adalah kemampuan suatu butir item tes hasil belajar untuk dapat membedakan antara *testee* yang berkemampuan tinggi dengan *testee* yang kemampuannya rendah demikian rupa sehingga sebagian besar *testee* yang memiliki kemampuan yang tinggi untuk menjawab butir item tersebut lebih banyak menjawab butir item tersebut lebih banyak yang menjawab betul, sementara *testee* yang kemampuannya rendah untuk menjawab butir item tersebut sebagian besar tidak dapat menjawab item dengan betul.

Rumus yang digunakan untuk mencari indeks daya pembeda adalah:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B}$$

Keterangan:

D = Indeks daya pembeda

JA = Banyaknya peserta kelompok atas

JB = Banyaknya peserta kelompok bawah

BA = Banyaknya peserta kelompok atas yang menjawab soal dengan benar

BB = Banyaknya peserta kelompok bawah yang menjawab soal dengan benar

(Suherman 2003: 159)

Adapun klasifikasi daya pembeda berdasarkan Suherman (2003: 161), dapat

dilihat pada Tabel berikut:

Tabel 4. Klasifikasi Daya Pembeda

Besar Daya Pembeda	Interpretasi
$DP \leq 0$	Soal sangat jelek
$0,00 < DP \leq 0,20$	Soal jelek
$0,20 < DP \leq 0,40$	Soal cukup
$0,40 < DP \leq 0,70$	Soal baik
$0,70 < DP \leq 1,00$	Soal sangat baik

(Suherman 2003: 161)

#### d. Tingkat Kesukaran Soal

Tingkat kesukaran adalah angka yang menunjukkan proporsi siswa yang menjawab betul suatu soal. Makin besar tingkat kesukaran berarti soal itu makin mudah demikian juga sebaliknya yaitu makin rendah tingkat kesukaran berarti soal itu makin sukar. Untuk menghitung tingkat kesukaran tiap butir soal digunakan persamaan:

$$P = \frac{B}{J_x}$$

Keterangan:

$P$  = indeks kesukaran,

$B$  = banyaknya siswa yang menjawab soal dengan benar, dan

$J_x$  = jumlah seluruh siswa peserta tes.

(Suherman, 2003: 146)

Adapun klasifikasi indeks kesukaran berdasarkan Suherman (2003:170) dapat dilihat pada Tabel berikut:

Tabel 5. Klasifikasi Indeks Kesukaran

Indeks Kesukaran	Interpretasi
IK = 0,00	Soal terlalu sukar
$0,00 < IK \leq 0,30$	Soal sukar
$0,30 < IK \leq 0,70$	Soal sedang
$0,70 < IK \leq 1,00$	Soal mudah
IK = 1,00	Soal terlalu mudah

(Suherman 2003: 170)

#### e. Efektifitas distraktor/pengecoh

Distraktor yaitu suatu pola yang dapat menggambarkan bagaimana *testee* menentukan pilihan jawabannya terhadap kemungkinan-kemungkinan jawab yang telah dipasangkan pada setiap butir item. Distraktor dinyatakan telah dapat menjalankan fungsinya dengan baik apabila distraktor tersebut sekurang-kurangnya sudah dipilih oleh 5% dari seluruh peserta tes atau

apabila mempunyai daya tarik yang besar bagi pengikut-pengikut tes yang kurang memahami konsep atau kurang menguasai bahan (kelompok bawah). Untuk menentukan angka pedoman efektifitas distraktor dapat dilakukan dengan cara perkalian antara  $5\% \times$  Jumlah siswa.

#### **E. Prosedur Analisis Soal**

Dalam sebuah penulisan soal agar mendapat soal yang berkualitas maka harus melalui beberapa tahap, yaitu:

##### **a. Analisis Secara Kualitatif**

Soal-soal mentah yang dibuat oleh para penulis soal berdasarkan kisi-kisi yang disusun Puspendik akan masuk dalam kategori soal mentah. Soal mentah akan ditelaah secara kualitatif sehingga diperoleh soal baik tanpa revisi dan soal yang perlu direvisi serta soal yang ditolak. Soal yang perlu direvisi akan langsung direvisi sehingga diperoleh soal yang baik dan soal yang ditolak akan dikembalikan ke penulis soal.

##### **b. Analisis Secara Kuantitatif**

Soal-soal yang baik secara kualitatif akan dirakit untuk proses ujicoba sehingga diperoleh data-data respon jawaban siswa. Respon jawaban siswa ini dianalisis menggunakan komputer sehingga diperoleh soal-soal yang baik dengan data-data parameter tingkat kesukaran dan daya beda untuk setiap butir soal.

## **BAB V**

### **SIMPULAN DAN SARAN**

#### **A. Simpulan**

Berdasarkan hasil penelitian dan pembahasan analisis soal Butir soal IPA Siswa Kelas IX MTs Guppi 03 Belanga Lampung Selatan Semester Ganjil Tahun Pelajaran 2015/2016 Kelas IX di MTs Guppi 03 Belanga Lampung Selatan maka dapat disimpulkan bahwa:

1. Soal Ulangan Tengah Semester Ganjil IPA Kelas IX di MTs Guppi 03 Belanga Lampung Selatan memiliki validitas logis, karena sudah sesuai dengan soal standar tetapi perlu perbaikan aspek konstruksi pada beberapa soal.
2. Butir soal IPA Siswa Kelas IX MTs Guppi 03 Belanga Lampung Selatan Semester Ganjil Tahun Pelajaran 2015/2016 reliabel dengan kategori tinggi.
3. Butir soal IPA Siswa Kelas IX MTs Guppi 03 Belanga Lampung Selatan Semester Ganjil Tahun Pelajaran 2015/2016 memiliki tingkat kesukaran sedang.
4. Butir soal IPA Siswa Kelas IX MTs Guppi 03 Belanga Lampung Selatan Semester Ganjil Tahun Pelajaran 2015/2016 memiliki daya beda baik.
5. Butir soal IPA Siswa Kelas IX MTs Guppi 03 Belanga Lampung Selatan Semester Ganjil Tahun Pelajaran 2015/2016 memiliki efektifitas pengecoh berfungsi.

6. Butir soal IPA Siswa Kelas IX MTs Guppi 03 Belanga Lampung Selatan Semester Ganjil Tahun Pelajaran 2015/2016 memiliki kualitas sesuai standar.

## **B. Saran**

Setelah dilakukan analisis butir soal IPA Siswa Kelas IX MTs Guppi 03 Belanga Lampung Selatan Semester Ganjil Tahun Pelajaran 2015/2016 maka disarankan agar sebaiknya Guru IPA MTs Guppi 03 Belanga Lampung Selatan melakukan uji coba dan analisis soal serta mempertahankan penyusunan soal yang berkualitas dengan berpedoman pada langkah pengembangan soal sesuai standar.

## DAFTAR PUSTAKA

- Afiyana, I. 2010. *Analisis kualitas soal latihan UN mata pelajaran IPA menggunakan program ITEMAN di batang (Skripsi)*. Universitas Negeri Semarang. Semarang.
- Afrianto, E. 2008. *Pengawasan Mutu Bahan/Produk Pangan Jilid II*. Jakarta: Departemen Pendidikan Nasional.
- Ariani, N.M. 2006. Karakteristik Dan Nilai Fungsi Informasi Butir Soal Ujian Akhir Sekolah Dasar Di Kota Bengkulu. *Jurnal Penelitian dan Evaluasi Pendidikan 2 (VIII)*: 213-230
- Arikunto, S. 2010. *Dasar Evaluasi Pendidikan*. Rineka Cipta. Jakarta.
- , 2013. *Prosedur Penelitaia Suatu Pendekatan Praktik*. Rineka Cipta. Jakarta.
- Ary. D., C.J Luchy dan R. Ashgar. 1982. *Pengantar Penelitian dalam Pendidikan* diterjemahkan oleh Arief Furchan 2007. Pustaka Pelajar. Yogyakarta.
- Azwar, S. 1987. *Tes Prestasi*. Liberty. Yogyakarta
- Basri, M. 2011. *Dasar-Dasar dan Perancangan Evaluasi Pembelajaran*. Unila. Bandar lampung
- Daryanto. 1999. *Evaluasi Pendidikan*. Rineka Cipta. Yogyakarta
- Handayani, A.P. 2014. *Pengembangan Tes Instrumen Pilihn Ganda Distraktor Bermakna Untuk Karakteristik Konsepsi Fisika Siswa Kelas XI SMA N 2 Malang (Skripsi)*. Universitas Negri Malang. Malang
- Hasan, A. 2008. *Marketing*. Media Pressindo. Jogjakarta
- Joesman. 2008, *Simulasi Jaringan Berbasis Paket Dengan Mempergunakan Simulator OPNET (Online)*, (<http://joesman.wordpress.com/-page/2/>), diakses tanggal 4 Desember 2015).

- Mardapi, D. 2008. *Teknik Penyusunan Instrumen Tes dan Non Tes*. Mitra Cendikia Press. Yogyakarta.
- Nurgiyantoro, B. 2001. *Penilaian Dalam Pembelajaran Bahasa dan Sastra*. Yogyakarta BPFE. Yogyakarta.
- Kancana, N. 1983. *Evaluasi Pendidikan*. Usaha Nasional. Surabaya
- Purnomo, A. 2007. Kemampuan Guru Dalam Merancang Tes Berbentuk Pilihan Ganda Pada Mata Pelajaran IPS Untuk Ujian Akhir Sekolah (UAS). *Jurnal Lembaran Ilmu Pendidikan* 36 (1): 1-6.
- Shakil, M. 2008. Does Brand Extension Impact Parent Brand: A Case of Johnson. *Journal UK*. Vol. 9, No. 1, Hal 53-66. (Online), (<http://www.researchgate.net>, diakses tanggal 2 Desember 2015)
- Sudjana. 2005. *Penilaian Hasil Proses Belajar Mengajar*. Remaja Rosdakarya. Bandung.
- Sudijono, A. 2006. *Pengantar Statistik Pendidikan*. PT. Raja Grafindo Persada. Jakarta.
- Suherman. 2003. *Evaluasi Pembelajaran Matematika*. JICA UPI. Bandung
- Surapranata, S. 2004. *Panduan Penulisan Tes Tertulis Implementasi Kurikulum 2004*. Remaja Rosdakarya. Bandung.
- Suryabrata, S. 2004. *Metodologi Penelitian*. Raja Grafindo persada. Jakarta.
- Widoyoko. 2009. *Evaluasi Program Pembelajaran*. Pustaka Pelajar. Yogyakarta
- Widodo. 2010. Analisis Butir Soal Tes. *Jurnal Pendidikan Penabur* No.14/Tahun ke-9/Juni 2010.
- Zaman dan Eliyawati. 2010. *Bahan Ajar Pendidikan Profesi Guru, Media Pembelajaran Anak Usia Dini*. Bandung: UPI.
- Zulkarnain. 2008. *Transformasi Nilai- Nilai Islam*. Pustaka Pelajar. Yogyakarta.